Modeling Transverse Orbit Feedback Control

Charles Schwartz

September 25, 2000

1 Introduction

There are two objectives to this report. The first is to present a model, or mathematical representation, of transverse orbit feedback control in general. The second objective is to utilize this model as an analysis tool for the APS real-time feedback system (RTFS). A most important example is that the model quantitatively predicts the effect of adding more correctors to the RTFS. This and other results are presented below in some detail.

The next section reviews orbit correction as practiced at the APS. In Section 3 the model is presented, thus fulfilling the first objective of the report. That section also lays the groundwork for APS-specific applications by describing how the model is parameterized for the APS. Section 4 contains applications of the model to the APS RTFS, and Section 5 is a summary of the report. Section 8 has two appendices: the first gives some model data, and the second appendix explores some properties of the RTFS, as revealed by the model.

In the report both elementary matrix operations and standard analysis tools for *discrete-time* signals and systems are used. Examples of the latter include the $\mathcal{Z}$-transform ($\mathcal{Z}$T), the Fourier transform (FT) (the two are used here interchangeably), and the bilinear continuous-time-to-discrete-time mapping. This basic material can be found in many texts; an example is [1]. The report keeps to the notational conventions of these disciplines as follows. Vectors and matrices are exclusively boldfaced, and exclusively lower-case and upper-case, respectively. Time dependence of scalar, vector, and matrix functions is denoted exclusively by square brackets, and frequency-dependence is denoted exclusively by parenthesis. Thus e.g. $\mathbf{x}[k]$ is a vector time-signal, and $\mathbf{x}(e^{j\omega})$ denotes its (element-wise) Fourier transform. Nonitalic subscripts are employed to label scalars, vectors, and matrices, while an italic subscript denotes a variable index. For example, if the set of

corrector error signals is denoted by $\mathbf{c}_{\mathrm{E}}[k]$, then its nth element may be denoted by $c_{\mathrm{E},n}[k]$. Finally, *scalar* frequency functions are capitalized but not scalar time functions; thus the nth diagonal element of the diagonal matrix $\mathbf{H}(e^{j\omega})$ may be denoted $H_n(e^{j\omega})$.

2 Orbit Correction Overview

Both in theory and practice one can distinguish between two orbit correction algorithms: static orbit correction and dynamic orbit correction [2]. Indeed, at the APS two such systems operate in parallel, see [3] and [4]. Both systems utilize beam position monitors (BPMs) to sense the beam's path deviation in both the horizontal and vertical planes, and then use corrector magnets to steer the beam back to the desired orbit in each plane. Both systems are essentially discrete-time systems. The difference in bandwidth between the static and the dynamic systems is what distinguishes them, as well as their control objective.

The objective of the APS slow orbit correction system is to minimize the orbit deviation at over 300 BPM points in the storage ring. Measured BPM data is sent to a workstation-based algorithm that computes the amount in Amperes by which each of 80 correctors distributed about the ring are to be driven. The current-driven correctors impart a magnetic field to the vacuum chamber, and thus steer the beam as close as possible to the desired orbit. The processing time for one such cycle of BPM-measurement-to-corrector actuation is one half-second.

The APS RTFS executes the same cycle of BPM-measurement-to-corrector actuation, but with different parameters. The system currently utilizes 38 correctors and 160 BPMs, has dedicated DSP hardware, and cycles at every $T_s = 600\mu$s. Aside from having different parameters, its control objective is also different: the RTFS seeks not to minimize the DC orbit deviation at each BPM point, but rather to minimize the spatially-averaged RMS beam motion on some frequency interval. This last statment is made precise in the next section. At the APS, the static global system and the dynamic RTFS can operate in parallel because they are decoupled by first-order low-pass and high-pass filters, respectively. The filters insure that the low-frequency portion of the RTFS bandwidth does not overlap with the bandwidth of the global system.

3 A Model of Dynamic Orbit Feedback

The qualitative description of the previous section is now mathematically formulated. Specifically, a mathematical model that quantifies how well the feedback system achieves the control objective is presented. The presentation is better understood by referring throughout to the system block diagram given in Figure 1. The first portion of this section formulates a machine-aspecific mathematical model of orbit control, while the second portion discusses how this model is parameterized for the APS RTFS using an injected drive signal $\mathbf{s}[k]$. This signal is described below; assume for the first portion of this section that the injected signal is equal to zero.

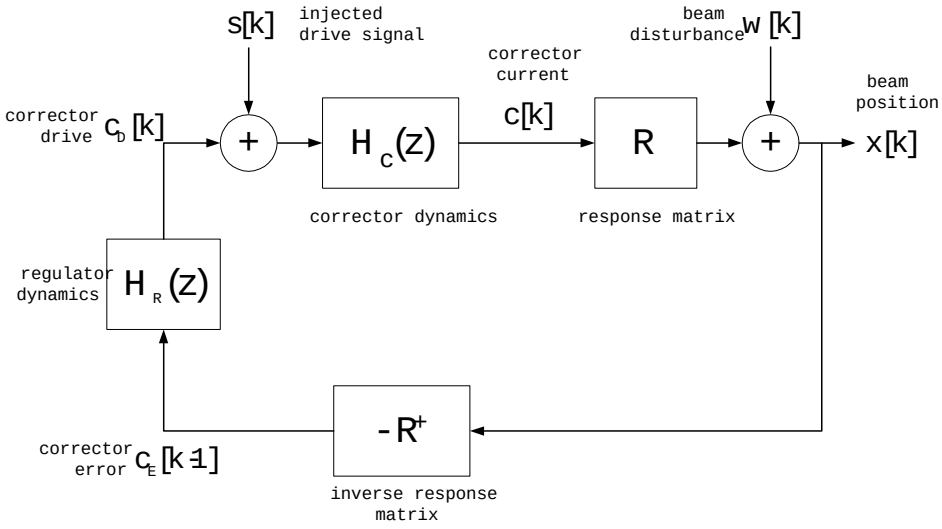

Figure 1: System block diagram.

3.1 A Mathematical Model

For each time sample k, let $\mathbf{x}[k]$ denote the column vector of orbit deviations as measured on the M BPMs. For the APS RTFS $\mathbf{x}[k]$ is a 160-by-1 vector, $M = 160$. This vector signal is processed by the feedback system to produce $N = 38$ corrector drive signals, which are collectively denoted by the N-tuple column-vector $\mathbf{c}_{\mathrm{D}}[k]$. As stated above, the control objective is to minimize the spatially-averaged RMS motion over some frequency band Ω. In other words, the objective is to minimize the square root of the scalar function $1/M||\mathbf{x}(e^{j\omega})||_2^2$ for all $\omega \in \Omega$, where $\mathbf{x}(e^{j\omega})$ is the Fourier transform (FT) of $\mathbf{x}[k]$, and $1/M||\mathbf{x}(e^{j\omega})||_2^2$ is its spatially averaged power spectral density (PSD).

The beam motion as measured at the BPMs, $\mathbf{x}[k]$, is affected by disturbances, which are denoted by the M-tuple $\mathbf{w}[k]$, and by orbit correction via the corrector dipoles. Typical disturbances include ground vibrations and magnet field fluctuations due to noisy power supplies. The beam motion is written as

$$\mathbf{x}[k] = \mathbf{R}\mathbf{c}[k] + \mathbf{w}[k], \tag{1}$$

where $\mathbf{R}$ is the M-by-N response matrix, and $\mathbf{c}[k]$ is the corrector current. It is desired to find a parametric model in the form of a transfer-function-like matrix that relates the RTFS nominal input $\mathbf{w}(z)$ to the nominal *closed-loop* output $\mathbf{x}(z)$. That is, a matrix $\mathbf{G}(z)$ is sought that can predict the RTFS' response to any input, $\mathbf{x}(z) = \mathbf{G}(z)\mathbf{w}(z)$.

In order to convert the orbit motion data $\mathbf{x}[k]$ into corrector-specific information, the RTFS computes the so-called corrector error $\mathbf{c}_{\mathrm{E}}[k]$, defined by

$$\mathbf{c}_{\mathrm{E}}[k] := -\mathbf{R}^{+}\mathbf{x}[k-1], \tag{2}$$

where $\mathbf{R}^{+}$ is the *pseudo*-inverse of $\mathbf{R}$ (see section 8.2, [2], [5]). Note from (2) that the corrector error is always delayed from the instantaneous $\mathbf{x}[k]$ by one tick, due primarily to a computation-time delay. The corrector current $\mathbf{c}(z)$ is simply given by

$$\mathbf{c}(z) = z^{-1}\mathbf{H}_{\mathrm{C}}(z)\mathbf{H}_{\mathrm{R}}(z)\mathbf{c}_{\mathrm{E}}(z), \tag{3}$$

where the *diagonal* N-by-N matrices $\mathbf{H}_{\mathrm{C}}(z)$ and $\mathbf{H}_{\mathrm{R}}(z)$ contain linear, time-invariant (LTI) discrete time system models of each nth corrector and regulator, respectively. Note that aside from the regulator dynamics, $\mathbf{H}_{\mathrm{R}}(z)$ also contains the effect of the bandpass filter that decouples the RTFS from the static global orbit feedback system. Specifically, $H_{\mathrm{R}}(z)$ equals the regulator $\mathcal{Z}$T multiplied by the

decoupling filter's $\mathcal{Z}$T.

The System Transfer Function Matrix Upon combining equations (1), (2), and (3), the transfer function matrix (TFM) $\mathbf{G}(z)$ arises, where $\mathbf{x}(z) = \mathbf{G}(z)\mathbf{w}(z)$, and

$$\mathbf{G}(z) = \left[\mathbf{I} + \mathbf{R}\mathbf{H}(z)\mathbf{R}^{+}\right]^{-1}. \tag{4}$$

The system dynamics are represented by the matrix $\mathbf{H}(z) := z^{-1}\mathbf{H}_{\mathrm{R}}(z)\mathbf{H}_{\mathrm{C}}(z)$. The TFM quantifies the effect of the RTFS in rejecting disturbances.

Recalling that the control objective is, in essence, minimization of the spatially averaged beam PSD, it is necessary to further develop the use of $\mathbf{G}(z)$ so that it answers to the control objective per se. To do this, note that

$$||\mathbf{x}(e^{j\omega})||_2^2 = \mathbf{x}^{*}(e^{j\omega})\mathbf{x}(e^{j\omega}), \tag{5}$$

where * denotes conjugate transpose (Hermitian). Thus follows the useful relationship

$$\frac{1}{M}||\mathbf{x}(e^{j\omega})||_2^2 = \frac{1}{M}\mathbf{w}'(e^{-j\omega})\mathbf{G}'(e^{-j\omega})\mathbf{G}(e^{j\omega})\mathbf{w}(e^{j\omega}), \tag{6}$$

where the superscript $'$ denotes transpose. Equation (6) relates the disturbance vector to the spatially averaged beam PSD as a linear function of the TFM.

There is one more step needed in order for the TFM to be a viable mathematical model. Specifically, equation (6) must be simplified further since it requires unavailable phase information for the disturbance vector $\mathbf{w}(e^{j\omega})$. The desired result is obtained by instead using the spatially averaged disturbance, denoted $w(e^{j\omega})$, which is defined via $w(e^{j\omega})\mathbf{i} = \mathbf{w}(e^{j\omega})$, where $\mathbf{i}$ is an M-tuple vector of ones. Under this replacement the following result is achieved:

AN ORBIT FEEDBACK MODEL *The spatially averaged (closed-loop) residual beam motion is related to the spatially averaged disturbance PSD and the RTFS transfer function matrix, according to*

$$\frac{1}{M}||\mathbf{x}(e^{j\omega})||_2^2 = \frac{1}{M}|w(e^{j\omega})|^2\mathbf{i}'\mathbf{G}'(e^{-j\omega})\mathbf{G}(e^{j\omega})\mathbf{i}, \tag{7}$$

where $\mathbf{G}(e^{j\omega}) = [\mathbf{I}+\mathbf{R}\mathbf{H}(e^{j\omega})\mathbf{R}^{+}]^{-1}$. This equation dictates how well the control objective is achieved by the RTFS. Moreover, under (7) the inherent spatiality of the RTFS is compressed, as the effect of the M^2-element TFM is reduced to a scalar. Depending on the application, the RTFS can thus be thought of as a scalar system or as spatially distributed system. An example of this dichotomy arises in model parameterization, which is discussed next.

3.2 Parameterization of the APS Real-Time Feedback System

To compute equation (7) one requires the spatially averaged disturbance PSD, $|w(e^{j\omega})|^2/M$, and a parameterized TFM $\mathbf{G}(e^{j\omega})$. The following addresses how this data can be determined, using the APS storage ring vertical motion as an example.

The Beam Disturbance Spectrum In open loop the beam motion as measured on the BPMs is equal to its disturbances there [see (1)]. Therefore the spatially averaged open-loop beam sample PSD tends to resemble the spatially averaged disturbance PSD, which is assumed to be stationary. That is,

$$\frac{1}{M}|w(e^{j\omega})|^2 = \frac{1}{M}\mathbf{x}^*(e^{j\omega})\mathbf{x}(e^{j\omega}) \qquad \text{in open loop, over time.} \tag{8}$$

In practice, only the salient frequency domain features of the often noisy disturbance spectrum are needed; therefore the disturbance PSD estimate is often interpolated and then filtered, thus yielding a smooth estimate of the beam disturbance. Open-loop PSD data and a smooth estimate of that data are plotted in Figure 2.

The Transfer Function Matrix In order to quantitatively determine $\mathbf{G}(z)$ it is necessary to determine both $\mathbf{R}$ and $\mathbf{H}_{\mathrm{C}}(z)$ [the regulator dynamics are known by design]. Measurement of the response matrix used in the RTFS is a routine exercise at the APS. It is accomplished by recording the BPM response in open loop to the N corrector drives, each driven one at a time. In terms of the variables defined above, this procedure corresponds to driving each of the N elements of the injected drive signal $\mathbf{s}[k]$ in open loop (thus $\mathbf{c}_{\mathrm{D}}[k] = \mathbf{0}$), and observing the steady state BPM values $\mathbf{x}[k]$ (see Figure 1).

Each sector in the storage ring contains usually eight correctors, numbered A1, ..., A4, B4, ..., B1. Currently only A3 correctors are used in the RTFS. It has been assumed to date that each type of corrector and associated power supplies have essentially identical dynamical properties across sectors (i.e., all A3's are similar, etc.). Indeed, recent data shows that all A3 correctors have roughly the same speed of response. However, there are significant differences, e.g. some A3 correctors overshoot and others do not. Still, it has been shown that a rough, average A3 corrector model is valid for assessing the overall RTFS performance, whereas assessment of spatially localized behavior requires that the corrector models be parameterized according to the individual correctors.

The standard method for parameterizing the linear model of a *scalar* dynamic system is to fit *in situ* input-output data to a transfer function. This scalar system

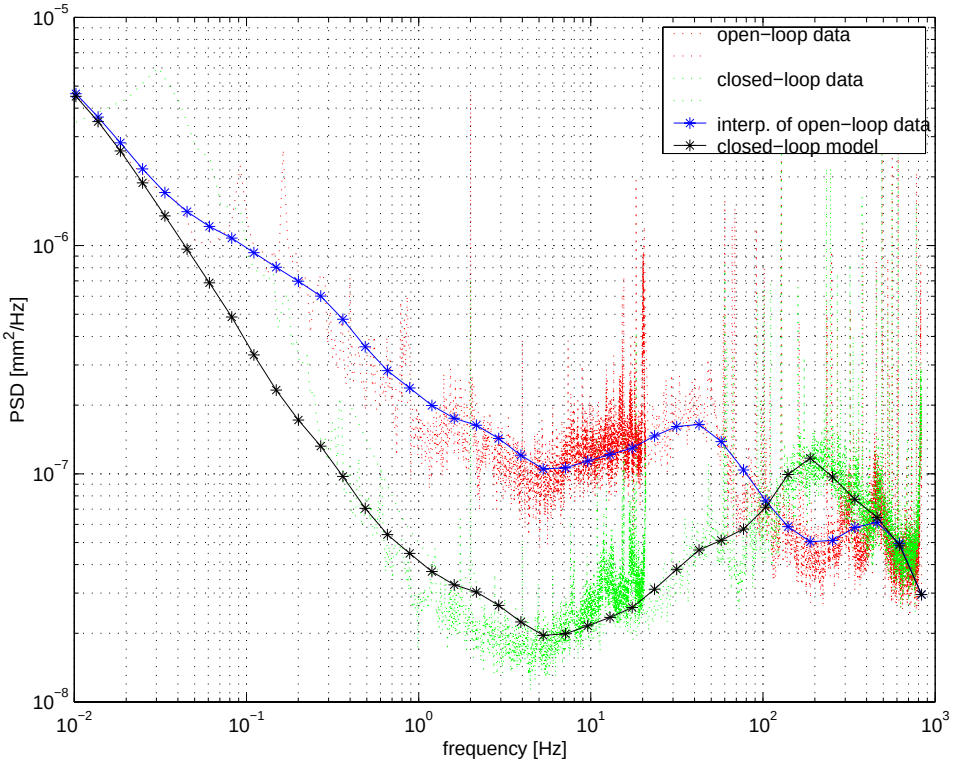

Figure 2: Open-loop spectrum and its interpolation; closed-loop spectrum as measured and predicted by the model.

method readily applies to the RTFS as well, since $\mathbf{R}^+$ decouples the N loops (see section 8.2). That is, *in situ*, i.e., with beam, and in open loop, exclusively one corrector is driven via the injected drive signal. If this input, e.g., a square wave, has a high signal-to-noise (i.e., $\mathbf{w}[k]$) ratio, then the corrector's effective response is the nth element of the corrector-error vector. In terms of the above notation, the corrector's input is the nth element of $\mathbf{s}[k]$, denoted $s_n[k]$, and the output is the nth element of the corrector-error vector, denoted $c_{\mathrm{E},n}[k]$. This input-output data pair is used to parameterize the corrector:

OPEN-LOOP PARAMETERIZATION PROCEDURE *Parameterization of each nth corrector is accomplished by fitting a transfer function to the high SNR, open-loop, input-output data $\{s_n[k], c_{\mathrm{E},n}[k]\}$, i.e.,*

$$c_{\mathrm{E},n}(z) = -z^{-1} H_{\mathrm{C},n}(z) s_n(z) \qquad \textit{open loop, high SNR,} \tag{9}$$

where $H_{\mathrm{C},n}(z)$ denotes the corrector's dynamics.

Moreover, it is possible to qualitatively verify the parameterization of each corrector. If $\mathbf{s}[k] = \mathbf{0}$, then in *open loop* the nth corrector error is simply the disturbance localized to that corrector, denoted by $o_n(e^{j\omega})$. [This is equal to the nth row of $\mathbf{R}^+$ multiplied by $\mathbf{w}(e^{j\omega})$.] Using a PSD estimate of this signal measured in open loop, one can test how well the empricial *closed-loop* corrector error PSD matches the one predicted by

$$|c_{\mathrm{E},n}(e^{j\omega})|^2 = \frac{|o_n(e^{j\omega})|^2}{|1 + e^{-j\omega} H_{\mathrm{R},n}(e^{j\omega}) H_{\mathrm{C},n}(e^{j\omega})|^2} \qquad \text{closed loop.} \qquad (10)$$

The exercise associated with equation (10) is referred to as the closed-loop verification procedure.

Results obtained by applying the open loop parameterization procedure to the A3 corrector in sector 4, vertical plane (S4:AV3), are shown in Figure 3. A sixth-order model is necessary to characterize the somewhat unsual response features; the model is provided in section 8.1. Data from the closed-loop verification procedure is also shown in Figure 4; the empirical closed-loop PSD data matches well with the data obtained via equation (10) using the open-loop PSD. Not all corrector parameterizations are verified as well as S4:AV3. Indeed it was found that some iteration between identification (9) and verification (10) was necessary in the parameterization process, so that the modeled dynamics simultaneously fit measured data and are still robust. Note that the A1 and B1 corrector dynamics have simpler response waveforms, and require only a third-order model.

Despite the fact that the A3 correctors have different responses, it has been observed that the responses of the *regulated correctors* are actually quite similar. This means that individual corrector parameterizations are robust for the regulator currently in use (given in section 8.1). Furthermore, it means that it *is* possible to utilize a single average model $H_{\mathrm{C}}(z)$ for each of the A3 correctors, an approach that coincides with the scalar interpretation of the RTFS system mentioned above. Figure 2 shows the results of (7) using as input the spatially-averaged disturbance PSD estimate (also shown in the Figure), and an RTFS TFM parameterized by a *common* low-order A3 corrector model (given in section 8.1). As was stated above, a common, single model for all correctors works well for spatially averaged data, but individual and generally unique corrector models are required for modeling localized behavior.

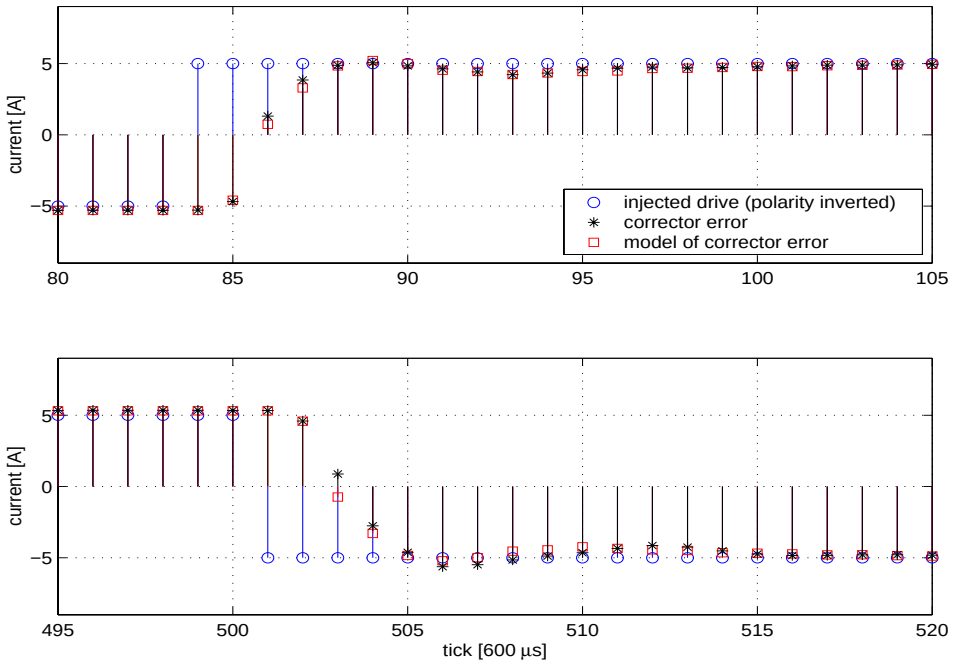

Figure 3: Open-loop parameterization of S4:AV3: corrector drive, and empirical and modeled corrector-error signals.

4 Applications of the Model to the APS RTFS

4.1 Expanding the RTFS to Include Additional Correctors

Intuitively, increasing the number of correctors in the RTFS should increase the amount of disturbance rejection, since there are now more spatial points of correction. However the TFM approach reveals that the improvement in disturbance rejection is frequency dependent – at high-frequencies the additional correctors may even induce disturbance *amplification*. The physical reason for this phenomenon lies in the bandlimited nature of the corrector-regulator chain: *in closed loop* low and middle frequencies of each corrector-error are rejected, but high frequencies are amplified (cf. Figure 2). This twin effect of low-frequency rejection and high-frequency amplification is borne out in the beam motion as well, and, assuming similar corrector dynamics, its amount scales with N (section 8.2). The situation is further complicated if the added corrector's dynamics are significantly different than those in the original setup. In this case the only way to quantitatively assess the effect of the additional correctors is to evaluate equation (7) for the original and

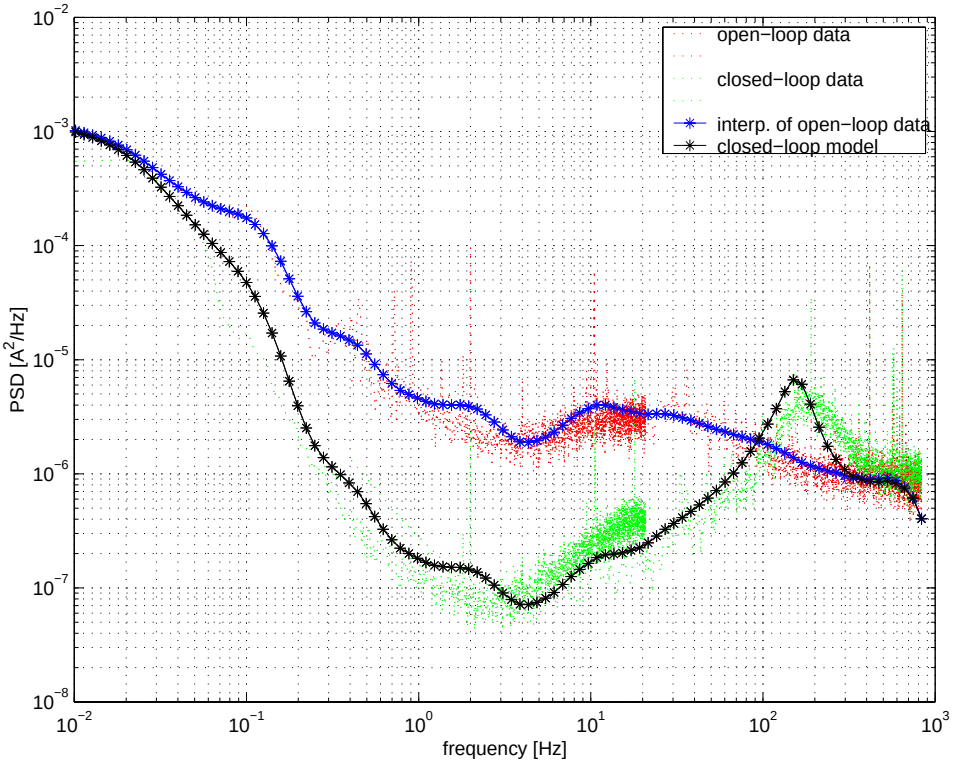

Figure 4: Closed-loop verification procedure applied to S4:AV3: open- and closed-loop corrector-error PSDs.

modified systems.

The question of adding correctors must be answered for the APS RTFS, since it is desired to further minimize the residual (closed-loop) beam motion by including an additional 79 correctors in orbit feedback, namely all of the A1 and B1 correctors. With 117 total correctors, there would be essentially three correctors per sector in the storage ring, whereas presently there is only one corrector per sector. Due to eddy-current effects the A1 and B1 correctors have a much slower response time than the A3 correctors (see [4] for details). This limits the bandwith of the modified RTFS, as discussed above. It is therefore desired to speed up the slow-corrector response times, a task that is accomplished by high-pass filtering the corrector drive signal. Design details of this "compensator" are deferred for a later subsection. For the present, it is enough to know that this is a necessary component of the proposed upgrade to the APS RTFS.

It is desired to evaluate via the TFM the PSD and RMS beam motion with and without the additional compensated correctors. This data gives a quantitative answer to the question of performance improvement. It is instructive to compare these to the RTFS performance under 117 correctors without slow-corrector compensation, so that the effect of the compensation can be readily seen. Also the fictitious case of 117 fast correctors (i.e., each having nominal A3 dynamics) conveys a best-case performance bound. In summary there are five cases to consider:

(a) open-loop beam motion

(b) residual motion, 38 correctors

(c) residual motion, 117 correctors, without slow-corrector compensation

(d) residual motion, 117 correctors, with slow-corrector compensation

(e) residual motion with 117 A3 correctors (fictitious)

Note that case (a) is merely the disturbance spectrum, and cases (b-e) are evaluated via equation (7) using the data from case (a) as input and using the appropriately parameterized TFM. Parameter data for the correctors is provided in the Appendix section 8.1.

The results for the five cases are simultaneously presented in Figure 5. The figures show first of all, that including more correctors in the RTFS brings more midband attenuation but also more high-frequency amplification. Comparing cases (c) and (d), it is seen that the compensator is a necessary component in the 117-corrector configuration, for it greatly increases the beam motion attenuation. Case (e) shows that only little high-frequency amplification occurs if the dynamics of the added correctors are equally as fast as the original correctors.

The data shown in the figure allows for some quantitative predictions. The compensated 117-corrector RTFS [case (d)] is predicted to achieve a maximum factor of four improvement *in PSD* over the 38-corrector configuration. This predicted impovement decreases with increasing frequency, until just over 100 Hz, at which point the modified system tends to amplify the disturbance more than otherwise. In RMS terms, on the frequency band $\Omega = [0.01, 100]$ Hz the predicted RMS is 1.67 μm, which is an improvement of 0.6 μm over the present 38-corrector RTFS. However, if the band of interest is increased to 450 Hz, then the modified RTFS *worsens* the RMS motion relative to the original system by over one micrometer (7.2 vs. 6.1).

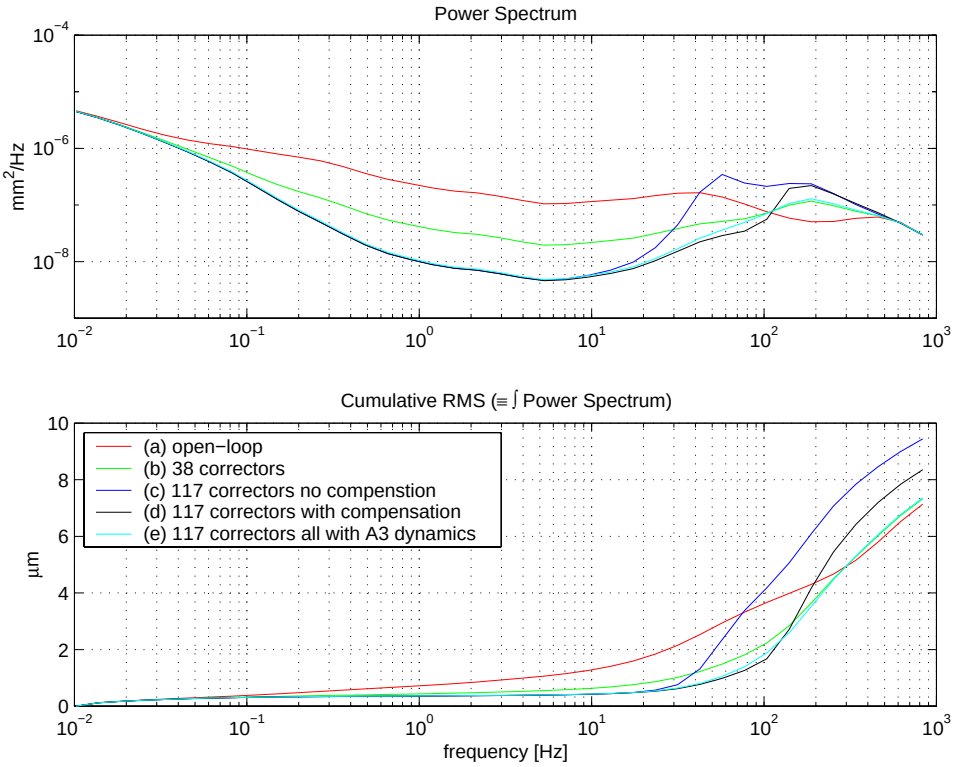

Figure 5: Beam motion in open loop and under various RTFS configurations.

In summary, the modified system yields a predicted RMS beam motion attenuation that, up until roughly 100 Hz, is approximately 25% smaller than with 38 correctors. However, the RMS up until 200 Hz and beyond is worse under the modification. Thus the upgrade is helpful only if a slightly higher level of high-frequency (above 200 Hz) beam motion is tolerable by the APS users. Further computations have shown that adjusting the regulator and compensator settings does not improve much upon these results. Indeed, *the principal limiting factor is the lattice parameters of the chosen correctors and BPMs and not the slower bandwidth of the A1 and B1 correctors* (as explained in Appendix section 8.2). This is borne out by the fictitious case (e): even if all of the correctors behaved as the fast A3 correctors, the system would still only achieve roughly *the same broadband RMS* improvement as with the compensated A1 and B1 correctors. In short, the slowness of the A1 and B1 correctors worsens the broadband RMS, but this worsening can be moved out to higher frequencies with compensation.

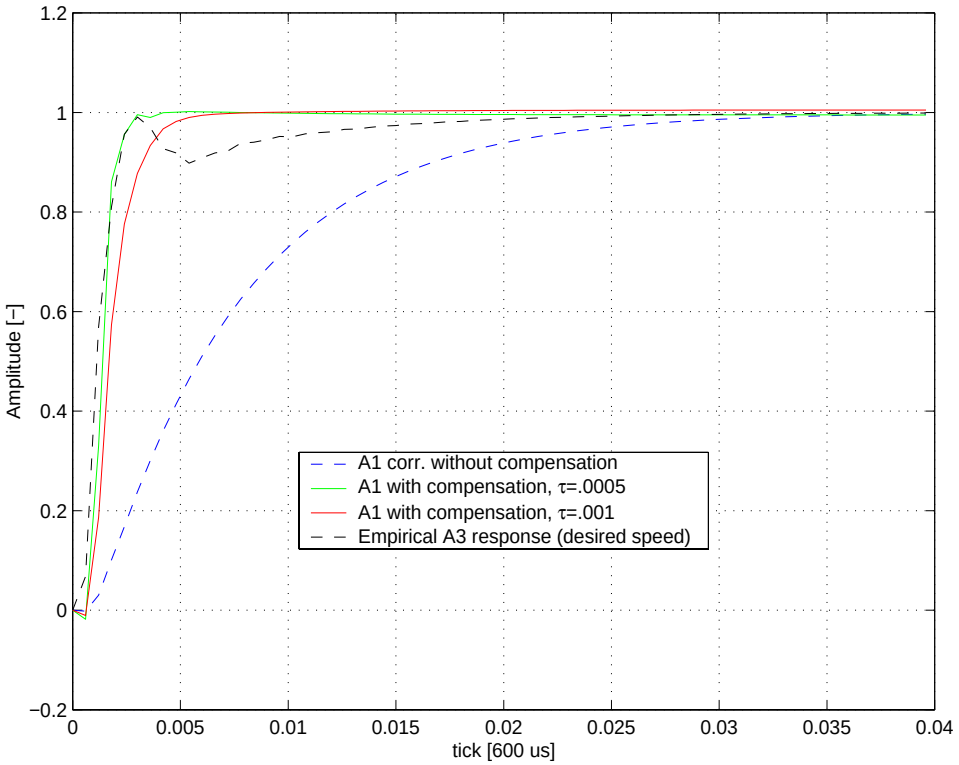

Figure 6: Step responses of compensated and uncompensated correctors.

Slow-Corrector Compensation The objective of the compensator design is to speed up the slow A1 and B1 corrector responses with a digital filter so that their speed of response approaches that of the fast A3 correctors. The compensator filter is designed by model matching: The corrector's (modeled) slow response is constrained through prefiltering to resemble some desired response. The desired response is the bilinear transform of a first-order system with time constant τ, cascaded with a two-tick delay. A system-identification algorithm is used to design the compensator, which emerges as a high-pass filter. Figure 6 shows how the compensator, as a function of τ, speeds up the A1 corrector's response, so that it no longer inhibits the RTFS as badly (the BV1 responses are roughly similar). The S4:AV3 response is included in the plot for reference. Of course, there are practical limits on the speed of the demanded corrector response, i.e. high-frequency gain contraints (see below), and robustness constraints. Specifically for the latter, if the compensator design is based on an overfitted corrector model, then the predicted increase in speed may not be practical, i.e., it may be oscillatory or closed-loop

unstable. Thus robustness considerations demand that the compensator design not seek a response that is too fast. The data shown in Figure 5 is with compensator parameter $\tau = 0.001$, so that the results reported above are realistically achievable.

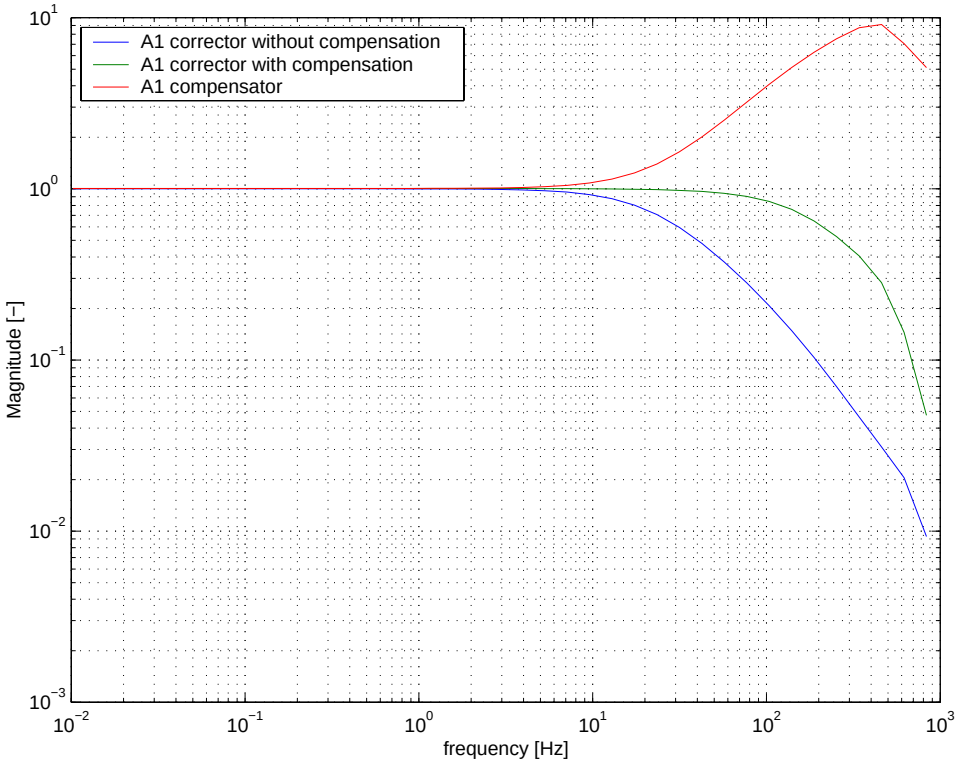

Figure 7: Frequency responses of compensated and uncompensated correctors and of compensator ($\tau = 0.001$).

Figure 7 gives frequency-response plots of the uncompensated and compensated A1 corrector, as well as the compensator itself. The compensator's high gain at high frequencies (nearly 20 dB) extends the bandwidth of the corrector. However, this plot does not take into account high-frequency gain constraints in the corrector's power supplies ("actuator effort"). Fortunately the corrector-error signals themselves are of small enough amplitude such that the demanded power supplies' output under compensation is achievable (cf. Figure 4): Figure 8 shows the cumulative RMS of the corrector drive with and without compensation, due to a typical open-loop corrector error. The figure shows that the compensator demands higher level power supply output at higher frequencies, and that the average

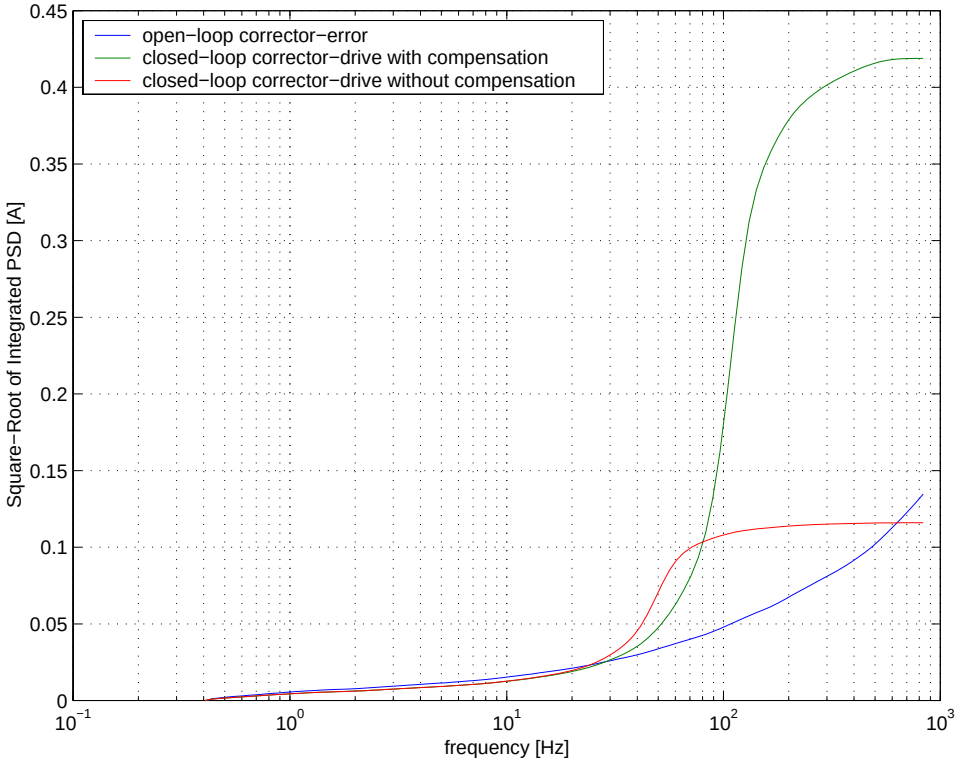

Figure 8: Cumulative RMS of compensated and uncompensated corrector drive due to typical open-loop corrector error.

demanded output (on the corrector) is ± 0.19 A. Further simulations show that the maximum demanded current on a sample-by-sample basis is within ± 0.7 A for a typical loop, but loops fighting a lot of noise might need as much 1.6 A. In short, the proposed compensator's demanded output is within the APS corrector power supplies' capability.

4.2 Other Applications and Further Study

Loop Coupling and Tuning It is shown in the Appendix that the N RTFS loops are decoupled to first order and that the RTFS can be thought of as N separate systems that operate in parallel and with no dynamic cross coupling. Therefore, there is no inherent performance benefit in "equalizing" the dynamics of each loop to be *identical*. Indeed, only the gross response speed of the fast correctors,

and not all of its full response features, is desired from the slow correctors (see Figure 6). Tuning the regulator/compensator of each loop individually improves the overall RMS beam motion. A special case of this can be seen in cases (c) and (d) of Figure 5: the performance of the RTFS as a whole improves as the dynamics of the slow loops are improved. Formal methods to optimize the RTFS by suitable tuning of each loop's regulator parameters and/or structure are under study.

The RTFS effective decoupling is valid to first order. However, when $\mathbf{R}^+$ is inaccurate then second-order effects begin to show up. These include low-level ghosting effects of one channel's corrector drive impinging on another channel's corrector error. The TFM analysis presented in the Appendix shows that *this coupling is in fact dynamic*, as has been observed in practice. This counterintuitive behavior is currently under study.

Disturbance Source Identification There are many low-amplitude disturbances that contribute to the observed orbit motion at the APS. However strong local sources could be potentially eliminated if they could be identified. The RTFS can be used to do this [4]. By choosing one of the available RTFS vector signals as a diagnostic signal, e.g., $\mathbf{x}[k]$ or $\mathbf{c}_{\mathrm{E}}[k]$, it is possible to spatially locate the disturbance source, at least roughly. For example, a noisy quadrupole or non-RTFS-corrector power supply in a given sector will induce high signal levels on the corrector errors of that sector (and adjacent sectors), indicating a problem in the vicinity. By combining spatial information with time-frequency information, it is anticipated that strong sources can be characterized and located.

However, one cannot assume that the frequency content of the *closed-loop* diagnostic signal corresponds to the frequency content of the disturbance source: the disturbance rejection afforded by the RTFS filters the source's waveform, especially if it falls in the RTFS bandwidth. Thus to obtain useful frequency information, the frequency-domain action of the RTFS must be unscrambled from the diagnostic signal. Applying the inverse of the (appropriate) TFM to the chosen diagnostic signal accomplishes this. This operation yields an estimate of what the diagnostic signal would look like in open loop. The frequency content of that filtered signal does correspond to the frequency content of the source.

5 Summary and Conclusions

A parametric model for real-time orbit feedback (RTFS) has been presented. The model takes the form of a transfer-function matrix (TFM), and data and general methods for empirically parameterizing this matrix have been presented. To date the model has only been approximately parameterized, but this is sufficient for many applications. The TFM mathematically describes the dynamics of orbit feedback, and thus can be used to quantify the performance benefit in adding more correctors to the APS RTFS. In particular, it was seen that by increasing the number of correctors per sector from one to three, the RMS residual vertical beam motion (0.01 Hz to 100 Hz) is decreased by approximately 25%, but the RMS over a wider frequency range is *increased*. Since in practice the loop components are bandlimited, this closed-loop high-frequency amplification is unavoidable, although it is mitigated by prefiltering the corrector drive with a realizable high-pass compensator filter. The slowness of the A1 and B1 correctors is the source of the additional high-frequency amplification; the compensator cannot get rid of the amplification, it can just shift it in frequency space, subject to gain and closed-loop stability constraints. The model is useful in approaching other orbit-feedback-related problems, most notably regulator tuning, disturbance-source identification, and loop coupling, since the TFM captures all of the dynamics of the RTFS. Work is underway to fully parameterize the model and to use it in other orbit control and RTFS-related applications.

6 Acknowledgments

The author wishes to thank John Carwardine of APS for many fruitful technical discussions, and for reviewing this report in great detail. Two additional members of the APS staff were also of assistance: F. Lenkszus helped in obtaining some of the data, and L. Emery gave important technical insight.

7 References

[1] *Feedback Control of Dynamic Systems*, G. F. Franklin, J. D. Powell, and A. Emami-Naeini (1994).

[2] *Handbook of Accelerator Physics and Engineering*, eds. A. W. Chao and M. Tigner, Ch. 4 (1999).

[3] "Advances in Orbit Drift Correction in the Advanced Photon Source Storage Ring," L. Emery and M. Borland, *Proc. 1997 Particle Accelerator Conference* (1997).

[4] "Real-Time Orbit Feedback at the APS," J. A. Carwardine and F. R. Lenkszus, *Proc. Eighth Beam Instrumentation Workshop* (1998).

[5] *Matrix Analysis*, R. A. Horn and C. R. Johnson (1985).

8 Appendix

8.1 Transfer Functions

Some of the transfer functions (TFs) mentioned in the text are:

$$\begin{aligned}
\text{S4:AV3 corrector TF} &= \frac{-0.07z^5 - 0.52z^4 - 0.12z^3 + 0.31z^2 + 0.31z - 0.04}{z^6 - 0.16z^5 - 0.76z^4 - 0.31z^3 + 0.50z^2 - 0.001z - 0.15} \\
\text{General A1 TF} &= \frac{0.001z^2 - 0.03z - 0.04}{z^3 - 0.78z^2 - 0.34z + 0.19} \\
\text{General A3 TF} &= \frac{-0.43z - 0.43}{z^2 - 0.2z} \\
H_{\mathrm{R}}(z) &= \frac{0.18z^2 - 0.18}{z^2 - 1.91z + 0.91} \qquad \text{(regulator TF)} \\
\text{A1 Compensator TF} &= \frac{6.07z^3 - 3.3z^2 - 2.03z}{z^3 - 0.30z^2 + 0.08z - 0.07}
\end{aligned}$$

8.2 The Role of $\mathbf{R}^+$

The success of the orbit control endeavor hinges largely on the accuracy of $\mathbf{R}^+$. More specifically, the matrix products $\mathbf{R}\mathbf{R}^+$ and $\mathbf{R}^+\mathbf{R}$ respectively determine (a) how well the RTFS rejects disturbances, and (b) whether any of the N feedback loops cross couple (i.e., whether with zero input one corrector's output affects another in closed loop). In fact, these two notions are one and the same. These statements are proven in the following.

The matrix $\mathbf{R}^+$ is given by $(\mathbf{R}'\mathbf{R})^{-1}\mathbf{R}'$ [5]. Thus,

$$\begin{aligned}
\mathbf{R}^+\mathbf{R} &= (\mathbf{R}'\mathbf{R})^{-1}\mathbf{R}'\mathbf{R} && (11) \\
&= \mathbf{I}. && (12)
\end{aligned}$$

The matrix product on the left-hand side of (11) is of interest because it arises in the TFM that relates $\mathbf{s}[k]$ and $\mathbf{c}_{\mathrm{E}}[k]$, chosen as the respective TFM input and output ($\mathbf{w}[k] \equiv \mathbf{0}$). This TFM $\mathbf{G}_1(z)$ is given by [cf. (10)]

$$\begin{aligned} \mathbf{c}_{\mathrm{E}}(z) &= -z^{-1}\left[\mathbf{I} + \mathbf{R}^{+}\mathbf{R}\mathbf{H}(z)\right]^{-1}\mathbf{R}^{+}\mathbf{R}\mathbf{H}_{\mathrm{C}}(z)\mathbf{s}(z) \qquad (13) \\ &= -z^{-1}\left[\mathbf{I} + \mathbf{H}(z)\right]^{-1}\mathbf{H}_{\mathrm{C}}(z)\mathbf{s}(z) \\ &=: -z^{-1}\mathbf{G}_1(z)\mathbf{H}_{\mathrm{C}}(z)\mathbf{s}(z). \qquad (14) \end{aligned}$$

[The symbol $=:$ means that the right-hand side equals the left-hand side by assignment, i.e., by definition.] Since $\mathbf{G}_1(z)$ and $\mathbf{H}_{\mathrm{C}}(z)$ are diagonal matrices, none of the N signals in $\mathbf{s}[k]$ cross couple with the N corrector errors. This means that the N loops are decoupled, hence the RTFS can be viewed as N parallel, uncoupled systems. This fact does not change even if $\mathbf{w}[k] \neq \mathbf{0}$ or if the dynamics of each loop are unique. An immediate consequence is that there is no inherent benefit in equalizing the loops to behave alike.

Due to modeling errors in $\mathbf{R}$, equation (12) holds only to first order, at best. Consequently the N loops are weakly coupled, per (13). This has been readily observed, e.g., in carrying out the Open-Loop Parameterization Procedure (9): when only one corrector is driven in open loop by an external signal, low signal-level "ghosts" of that signal are observed on other corrector-error channels. While this static coupling is an intuitive result of $\mathbf{R}^{+}\mathbf{R} \neq \mathbf{I}$, it is perhaps not intuitive that in closed loop the ghosting effect is in fact frequency dependent, i.e., dynamic.

To better see how, e.g., low levels of this dynamic coupling arise, assume for simplicity that $\mathbf{H}(z) = H(z)\mathbf{I}$, and let $\mathbf{E} := \mathbf{R}^{+}\mathbf{R} - \mathbf{I}$ define an error matrix. Then suppressing z-dependence of $H(z)$, $\mathbf{G}_1(z)$ can be rewritten as

$$\mathbf{G}_1(z) = \left[(1+H)\mathbf{I} + H\mathbf{E}\right]^{-1}. \qquad (15)$$

Assume that $|||\frac{-H}{1+H}\mathbf{E}||| < 1$, where $|||\cdot|||$ is a matrix norm. This last condition is usually satisfied if the error matrix is "small." Then using the lemma [5]

$$(\mathbf{I} - \mathbf{C})^{-1} = \sum_{k=0}^{\infty}\mathbf{C}^k \qquad \text{if} \quad |||\mathbf{C}||| < 1, \qquad (16)$$

it can be shown that

$$\mathbf{G}_1(z) = \frac{1}{1+H}\sum_{k=0}^{\infty}\left(\frac{-H}{1+H}\right)^k\mathbf{E}^k. \qquad (17)$$

Thus there are $(\frac{-H}{1+H})^k$-filtered off-diagonal terms in $\mathbf{G}_1(z)$, whose magnitudes are dictated by $\mathbf{E}^k$, and these induce dynamic ghosting. Note however that (17) holds only for small errors associated with (12) [the lemma only specifies a sufficient condition]. If there are large errors, then the dynamic ghosting effect may have a different, unpredictable form. For example, a closed-loop $\frac{1}{(1+H)^2}$-filtering effect has been observed using a response-matrix model that does not satisfy $|||\frac{-H}{1+H}\mathbf{E}||| < 1$.

Finally, while a perfect $\mathbf{R}^+$ satisfies (12) and hence decouples the loops, the product $\mathbf{R}\mathbf{R}^+$ is only as "close as possible" to $\mathbf{I}$ [namely, in a Euclidian-norm (l_2) sense]. Indeed, the closer it is to $\mathbf{I}$ the better the disturbance rejection. To better see this, rewrite the TFM $\mathbf{G}(z)$ using the matrix-inversion lemma [5]:

$$\begin{aligned}\mathbf{G}(z) &= \mathbf{I} - \mathbf{R}\mathbf{H}(z)\left[\mathbf{I} + \mathbf{R}^+\mathbf{R}\mathbf{H}(z)\right]^{-1}\mathbf{R}^+ \\ &= \mathbf{I} - \mathbf{R}\mathbf{H}(z)\mathbf{G}_1(z)\mathbf{R}^+, \end{aligned} \tag{18}$$

where the diagonal $\mathbf{G}_1(z)$ was defined in (14). Assume, without much loss of generality, that the N loops have identical dynamics, $\mathbf{H}(z) = H(z)\mathbf{I}$. Then, $\mathbf{G}(z)$ simplifies to

$$\mathbf{G}(z) = \mathbf{I} - \frac{H(z)}{1+H(z)}\mathbf{R}\mathbf{R}^+. \tag{19}$$

Thus the maximal disturbance rejection $[1/1 + H(z)]\mathbf{I}$ is achieved if $\mathbf{R}\mathbf{R}^+ = \mathbf{I}$. Therefore, *the lattice parameters of the RTFS correctors and BPMs ultimately dictate the amount of RTFS disturbance rejection* since they determine to what degree $\mathbf{R}\mathbf{R}^+ \to \mathbf{I}$. On the other hand, depending on $H(z)$, the ratio $1/1 + H(z)$ may also *amplify* high frequencies. Indeed, for the APS RTFS, as more correctors are added to orbit feedback, i.e., as $N \to M$, both the disturbance rejection at low frequencies *and* the disturbance amplification at high frequencies increase. The *spatially averaged* RTFS behavior is governed by $\hat{G}(z) = \frac{1+\epsilon H(z)}{1+H(z)}$, where

$$\epsilon := -\frac{1}{M}\sum_{m=1}^{M}\sum_{q=1}^{M} E_{m,q}, \tag{20}$$

and, as in (15), $[E_{m,q}] = \mathbf{E} := \mathbf{R}\mathbf{R}^+ - \mathbf{I}$. Finally, note that equations (18) and (19) are useful for optimal and loop-by-loop regulator design; this is currently under study.